DOCUMENT RESUME

ED 409 374 TM 026 945

AUTHOR McNeil, Keith
TITLE Directional and Non-directional Hypothesis Testing: A Survey of SIG Members, Journals, and Textbooks.
PUB DATE Mar 97
NOTE 12p.; Paper presented at the Annual Meeting of the American Educational Research Association (Chicago, IL, March 24-28, 1997).
PUB TYPE Reports - Evaluative (142) -- Speeches/Meeting Papers (150)
EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Editors; *Hypothesis Testing; *Scholarly Journals; *Statistical Significance; Surveys; *Test Use; *Textbook Content; Textbooks
IDENTIFIERS *Directionality

ABSTRACT

The use of directional and nondirectional hypothesis testing was examined from the perspectives of textbooks, journal articles, and members of editorial boards. Three widely used statistical texts were reviewed in terms of how directional and nondirectional tests of significance were presented. Texts reviewed were written by: (1) D. E. Hinkle, W. Wiersma, and S. G. Jurs (1994); (2) G. V. Glass and K. D. Hopkins (1996); and (3) R. C. Sprinthall (1990). All three focused on nondirectional tests of significance. While all three texts introduced one-tail (directional) tests of significant, two downplayed its significance. Research methods texts reviewed were by: (1) J. H. McMillan (1992); (2) R. E. Slavin (1992); and (3) R. C. Sprinthall, G. T. Schmutte, and L. Sirois (1991). Directional hypothesis testing was not given as much respect as nondirectional in any of the three texts. A review of 11 issues of "School Science and Mathematics" and 2 issues of the "Journal for Research in Mathematics Education" found nondirectional tests usually used, although the authors made directional conclusions. A survey completed by 10 members of the editorial board of "Multiple Linear Regression," members of the Special Interest Group (SIG), showed strong feelings by 3 board members that directional hypothesis testing should be thoroughly understood and used appropriately. Attachments present three text examples. (Contains three tables and nine references.) (SLD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

Directional and non-directional hypothesis testing: A survey of SIG members, journals, and textbooks

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY
Keith McNeil
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

Keith McNeil
O. D. Hadfield

New Mexico State University
AERA Chicago 1997

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
☑ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

ED 409 374

This present paper investigates the use of directional and nondirectional hypothesis testing from the three perspectives of (a) textbooks, (b) journal articles, and (c) editorial board members.

Textbooks

Statistical Texts

Three statistical texts were reviewed in terms of how directional and nondirectional tests of significance were presented. The focus was on (a) the fit between the research hypothesis (alternative hypothesis) and the test of significance, and (b) the interpretation of the test of significance.

Three of the more widely-used statistical texts were reviewed: (a) Hinkle, Wiersma, and Jurs (1994), (b) Glass and Hopkins (1996), and Sprinthall (1990). As seems reasonable, all three introduced hypothesis testing with testing a single population mean. The sole focus was on nondirectional tests of significance, probably to reduce the cognitive overload. All three texts discussed one-tail (directional) tests of significance, but two of them downplayed its importance. The directional test was the primary focus in only one text (Hinkle et al.). Sprinthall depicts a particularly gloomy role for the directional test of significance, "Because of the possibility of statistical slight of hand, some statisticians believe that the one-tail test should never be used. They insist that the only legitimate alternative hypothesis for the t test is the tried and true Mu1 = Mu2."

Sprinthall's (1990) first examples are nondirectional tests of a single population mean. These examples are all trying to "prove the null." In another early example, the scenario is presented that a communication will change political attitude--a directional expectation, but it is tested nondirectionally, and then interpreted directionally. Many other examples are presented in this same way (e.g., chi square, p. 293, chi square, p. 296, chi square, p. 298, chi square, p. 302, correlation, p. 304, p. 323, paired t test, p. 349, and experimental/control-pretest/posttest, p. 356). Sprinthall cautions the reader that many researchers have fudged their research hypothesis after looking at the data.

Glass and Hopkins (1996) give little attention to directional hypotheses. Indeed, they treat them as afterthoughts. Yet these authors make directional conclusions from

TM026945

BEST COPY AVAILABLE

tests of nondirectional hypotheses, even when indicating that directional hypotheses could have been tested. Almost all of their examples clearly call for a directional hypothesis (e.g., increase due to special treatment, practice effect, and school bond receiving a majority of the vote). The special treatment example actually resulted in a t value that was not significant with the nondirectional critical value, but would have been significant with, what we would argue, is the appropriate directional critical value (see p. 284). Also, in all of these texts, directional conclusions are made from nondirectional tests of significance, a conclusion specifically not allowed by the admission of each of the authors.

While all texts discuss the possibility of directional tests of significance when testing (a) difference between two means, (b) single proportion, (c) difference between two proportions, and (d) correlations, only one discussed directional tests with a priori planned comparisons, and none discussed directional tests with chi-square analysis, linear interaction, or beta weights in a multiple linear regression analysis. Yet directional conclusions were made by at least one of these texts in each of these areas. Of more relevance, directional hypotheses could have been tested in all these cases (see McNeil, 1996 for discussion of this issue). [It should be noted that Shavelson (1996) and McNeil, Newman, and Kelly (1996) are both particularly careful regarding introducing both directional and nondirectional testing for all of these tests of significance.

Table 1

Examples Where Author Made Directional Interpretation from Nondirectional Tests of Significance in Three Textbooks on Selected Statistical Tests

Selected Statistical Test	Author and page
Difference between two means	S 176 G 284 G 306
Diff between two dependent means	S 349 G 298 S 356
Single Population Proportion	G 328
Correlation	S 323 S 304
Difference between Two Correlations	G 360 G 363
Multiple comparisons	G 448
Chi Square	H 546 H 552 S 293 S 296 S 302 S 304

Note. S= page number from Sprinthall (1990)
G = page number from Glass and Hopkins (1996)
H = page number from Hinkle et al. (1994)

Educational Research Methods Texts

Although the major focus of these texts is not statistics, these texts do guide the

research methods of researchers, including the statistical reasoning. Furthermore, these texts are in a position to tie the research hypothesis with the research design and the research design with the statistical test of significance.

The three research methods texts reviewed were (a) McMillan (1992), Slavin (1992), and Sprinthall, Schmutte, and Sirois (1991). Each of the three texts covers statistical hypothesis testing in a cursory fashion, as that area is not the major focus of the text. But researchers are guided in their thinking by these texts, possibly more so than by a statistics text. Given that the research hypothesis and the research design guide the statistical analysis, one could argue that research methods texts are more crucial to understanding the logic of when to use a directional test of significance than are statistics texts. Unfortunately, directional hypothesis testing is not given as much respect as nondirectional in any of the three texts.

McMillan (1992) implies that only nondirectional conclusions can be made from a nondirectional test, yet presents three t test examples wherein two of them he makes a directional conclusion. He also makes a directional conclusion from an interaction plot, and a directional conclusion from a nondirectional 2*4 chi square test.

Sprinthall et al. (1991) discuss one-tail testing in detail, stating, " If, however, the hypothesis does specify the direction of the difference (that one set of sample scores will be higher than the other), the one-tail evaluation is usually preferred" p.139. We would take issue with the word "usually," as it seems only logical that if the hypothesis specifies direction that the statistical test must also reflect that direction. (Sprinthall et al. use the phrase "can be selected" later in the same paragraph, again not a strong endorsement for the directional test.)

Sprinthall et al. (1991) is the only reference reviewed that discusses various non-parametric techniques. It is interesting to note that most of the examples call for a directional test of significance, but are traditionally tested with a nondirectional distribution. The following nonparametric procedures are presented with directional examples, but nothing is said regarding a directional test of significance (a) 2*2 chi square, p.153, p. 154, (b) 2*2 change, p. 155, (c) Wilcoxon T test, p. 151, and (d) Mann-Whitney U test, p. 150.

Slavin (1992) takes a much stronger stand against the directional hypothesis testing. He says, "While use of a one-tail test makes it somewhat 'easier' to find significant differences between means, one-tailed tests (and directional hypotheses) are appropriate only when there is no conceivable chance that the results will turn out in a direction opposite to the one hypothesized" p. 156. This is a strange conclusion from the previous page where he stated, "In most experiments in which we are comparing two groups, we have a strong hypothesis to predict which group will have the higher mean. For example, if a researcher developed a new method of teaching geometry, she would have a hypothesis that the experimental group (which received the new method) would do better on the posttest than the control group (which studied geometry using traditional methods)" p. 155. Slavin is concerned that researchers need to allow for the possibility that the experimental students would learn less than the traditional students. But who would care if the new treatment is significantly worse than the traditional method? Given such an interest, almost anyone could devise a new procedure that is worse than the traditional procedure. So What?

Slavin makes a directional conclusion from a nondirectional t test (p. 164), uses a nondirectional test to test a directional hypothesis about a single population mean (p.165), uses a nondirectional ANCOVA to test a directional research hypothesis (p. 173), and makes a directional conclusion from a nondirectional test of a correlation coefficient (p. 180-182). While we would argue that these directional conclusions are not valid from a nondirectional test of significance, it is interesting to note what Slavin says about one-tailed hypotheses, "However, few if any relationships in social science are so well established and reliable that the possibility of a finding in the opposite direction to that hypothesized can be completely ignored. Hence, one-tailed tests should rarely be used in social science research" p. 156-157.

Journal Articles

In an attempt to present the actual application of directional hypothesis testing in the applied literature, we chose issues of *School Science and Mathematics* and *Journal for Research in Mathematics Education*. These journals were chosen because they do report some quantitative research and they were on the shelves of the second author. We do not purport that they are a random sample of anything, but we do assert that these issues are representative of the issues of those two journals. Given the confusing messages contained in the statistical and research methods texts, it was not surprising that the authors of the journal articles often present a confused picture. We reviewed 11 issues of *School Science and Mathematics* and 2 issues of the *Journal for Research in Mathematics Education*. About half of the articles contained tests of significance. In only one article did we find that a directional interpretation resulted from a directional test of significance that tested a directional hypothesis that emanated from the literature review. As indicated in Table 2, most tests of significance were nondirectional, although the author(s) made directional conclusions. In all cases, the literature review clearly implied a directional research hypothesis, although a nondirectional research hypothesis was stated or tested.

Table 2

Treatment of Directional Hypothesis Testing in 22 Journal Articles

Article #	RH in Lit Review	Directional RH specified	RH follows from Lit Review	Directional RH Tested	Directional RH Interpreted Correctly	Nondirectional RH Interpreted Correctly
1	Y	N	N	N	NA	N
2	Y	N	N	N	NA	N
3	Y	N	N	N	NA	N
4	Y	N	N	N	NA	N
5	Y	N	N	N	NA	N
6	Y	N	N	N	NA	N
7	Y	Y	Y	Y	Y	NA
8	Y	N	N	N	NA	N
9	Y	N	N	N	NA	N
10	Y	N	N	N	NA	N
11	Y	N	N	N	NA	N
12	Y	N	N	N	NA	N
13	Y	N	N	N	NA	N
14	Y	N	N	N	NA	N
15	Y	N	N	N	NA	N
16	Y	N	N	N	NA	N
17	Y	N	N	N	NA	N
18	Y	N	N	N	NA	N
19	Y	N	N	N	NA	N
20	Y	N	N	N	NA	N
21	Y	N	N	N	NA	N
22	Y	N	N	N	NA	N

Note. Y = Yes, attribute contained in the article.
N = No, attribute not contained in the article.
NA = Not Applicable, either the directional or nondirectional test of significance not conducted.

Editorial Board Members

A survey was distributed to all 10 of the current and previous editorial board members of the journal sponsored by the Special Interest Group: Multiple Linear Regression. Three scenarios were presented to the respondents. The first scenario indicated that two new and different teaching methods were to be compared. Respondents were given the option of three research hypotheses: (a) nondirectional, (b) directional in one direction, and (c) directional in the other direction. The four respondents all correctly chose the hypothesis of no difference, as there was no information on which to make a directional hypothesis. The sample data indicated that the two means were different, and one person concluded that the one population was higher. Two others said that the two populations were different.

The second scenario identified a current way of teaching and a new way of teaching. Three of the four respondents indicated that the research hypothesis would state that the new way should be better than the current way. The sample results were indeed not in this direction. One of these respondents said, "The new way is not an improvement" and another said, "The outcome is counter, and hence suggests the posited hypothesis was generated from a theory that is likely to be false."

The third scenario again identified a current way of teaching as well as a new way, but the sample results were in the hypothesized direction. The same three respondents identified the research hypothesis as the new method being greater than the current method.

In addition, the survey asked, "When would you use a directional test of significance? Check all that apply. Table 3 contains the results.

Table 3

Frequency of responses to the question, "When would you use a directional test of significance?

Frequency	Situation
3	Testing value of a new method as compared to the "tried and true method."
1	Checking on the reliability of a measure.
1	Checking on the validity of a measure.
3	Literature review supports such an hypothesis.
1	When using stepwise regression.
1	When using structural equation modeling.
1	Only when there is one degree of freedom in the numerator of the F test.
3	When the alternate hypothesis (or research hypothesis) calls for it.
1	Never.
1	(Supplied by respondent) When some combination of past research, authority, judgment, intuition, and common sense calls for it.

Finally, the survey posited the following question, "Why do some researchers make a directional conclusion from a nondirectional hypothesis?

One respondent said, "You've got me. It's illogical."

Another said, "Suppose Ho: rho = 0 and H1: rho $\neq$ 0, and you get r = .85. I think most people would conclude that there is evidence to support rho > 0."

A third respondent said, "Because they hope to capitalize on the results even though they didn't, when following classical hypothesis testing procedures, have the conviction or the inclination to take advantage of the additional power to use a directional hypothesis. Some users also may not notice if their computer package is

using a one- or two-tailed significance level with t tests. No guts, No glory with wimpy two-tailed test users."

The last respondent sums up our feeling and understanding of directional hypothesis testing: "My axiom on using directional hypotheses is that you should use them and only if you know what you're doing. If you don't know what you're doing, then why are you doing it?"

Summary

The review of the statistical and research methods textbooks indicated that many authors confuse the issue of directional and nondirectional tests of significance. Though there is a preponderance of desires to make directional interpretations, most authors encourage the student to make a nondirectional test of significance, and then make the directional interpretation.

Journal articles, at least in the two journals that we reviewed, follow this same pattern. In only one case did we find an article that posited a directional research hypothesis and then proceed to test it correctly--resulting in a correct directional interpretation. As was the case with the statistical authors, journal authors dearly want to make directional conclusions--and in most cases these directional conclusions could have been predicted before the research was undertaken.

What was somewhat encouraging was that three of the editorial board respondents took a strong stance on this issue. Hopefully more journal editorial board members will become aware of this issue and in the process improve the quality of research appearing in the journals.

References

Glass, G. V., & Hopkins, K. D. (1996). *Statistical methods in education and psychology* (3rd ed.). Needham Heights, MA: Allyn & Bacon.

Hinkle, D. E., Wiersma, W., & Jurs, S. G. (1994). *Applied statistics for the behavioral sciences* (3rd ed.). Boston: Houghton Mifflin Company.

McMillan, J. H. *Educational research: Fundamentals for the consumer.* New York: Harper Collins (1992)

McNeil, K. (1996). Testing directional research hypotheses. *Multiple Linear Regression Viewpoints, 21*(1), 23-31.

McNeil, K., Newman, I., & Kelly, F. J. (1996). *Testing research hypotheses using the general linear model.* Carbondale: Southern Illinois University Press.

Shavelson, R. J. (1996). *Statistical reasoning for the behavioral sciences* (3rd ed.). Boston: Allyn & Bacon.

Slavin, R. E. (1992). *Research methods in education* (2nd ed.). Boston: Allyn and Bacon.

Sprinthall, R. C. (1990). *Basic statistical analysis* (3rd ed.). Englewood Cliffs, NJ: Prentice Hall, Inc.

Sprinthall, R. C., Schmutte, G. T., & Sirois, L. (1991). *Understanding educational research.* Englewood Cliffs, NJ: Prentice Hall.

Carter, G., & Norwood, K.S. (1997). The relationship between teacher and student beliefs about mathematics. School Science and Mathematics, 97(2), 62-67.

Hypotheses

Student beliefs about mathematics are related to teacher beliefs about mathematics. Therefore, those in the reform classes will have higher scores on the NCTM Reform belief inventory. Series of 7 F tests, one for each subscale of the instrument; two groups.

1. Directional hypothesis indicated in literature review. Yes.

2. Directional hypothesis specified. No.

3. Hypothesis follows from review. No.

4. Directional hypothesis tested. No.

5. Directional hypothesis interpreted correctly. N.A.

6. Non-directional hypothesis interpreted correctly. No. Direction was claimed.

Two hypotheses would have been significant if $p = .09$ were to be divided by 2. These were two of 7 F-tests that were performed.

(5)

Wang, J., Wildman, L., & Calhoun, G. (1996). The relationship between parental influence and student achievement in seventh grade mathematics. School Science and Mathematics, 96(8), 395-399.

Hypotheses

Several parental factors (30) will be related to mathematics achievement. Multiple regression (backwards stepwise) was run, with all 30 variables.

1. Directional hypothesis indicated in literature review. Yes.

2. Directional hypothesis specified. No.

3. Hypothesis follows from review. No.

4. Directional hypothesis tested. No.

5. Directional hypothesis interpreted correctly. N.A.

6. Non-directional hypothesis interpreted correctly. No. Direction was claimed for all 12 relationships that were found. Actually, those that were in the wrong direction should be tossed out, and those in the hypothesized direction should have p divided by 2.

(4)

Boone, W.J. (1997). science attitudes of selected middle school students in China: A preliminary investigation of similarities and differences as a function of gender. School Science and Mathematics, 97(2), 96-103.

Hypotheses

There will be no significant differences between the male and female survey of science attitude questions.

23 different Chi-squares were run at $p < .05$, non-directional.

1. Directional hypothesis indicated in literature review. Yes.

2. Directional hypothesis specified. No.

3. Hypothesis follows from review. No.

4. Directional hypothesis tested. No.

5. Directional hypothesis interpreted correctly. N.A.

6. Non-directional hypothesis interpreted correctly. No. Direction was claimed for the various 8 Chi-squares that were significant at $p < .05$, although two-tailed was used initially. Some results were in the opposite direction to what should have been hypothesized, yet they were still directionalized after the fact.

In three cases, significance would have been found if probability had been divided by 2.